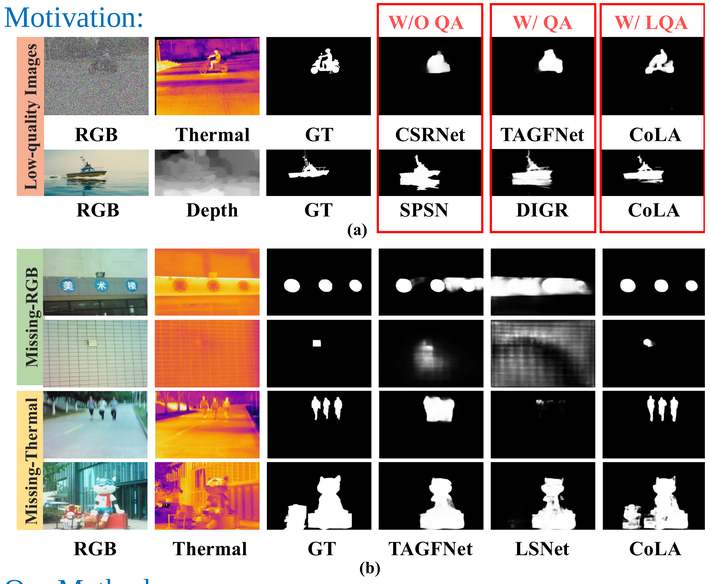
CoLA: Conditional Dropout and Language-driven Robust Dual-modal Salient Object Detection

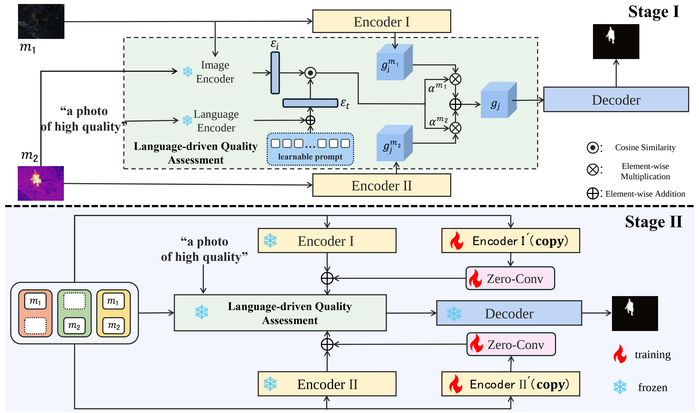
Shuang Hao, Chunlin Zhong, He Tang
{shuanghao, clzhong, hetang@hust.edu.cn}

EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

Motivation:



Our Method:

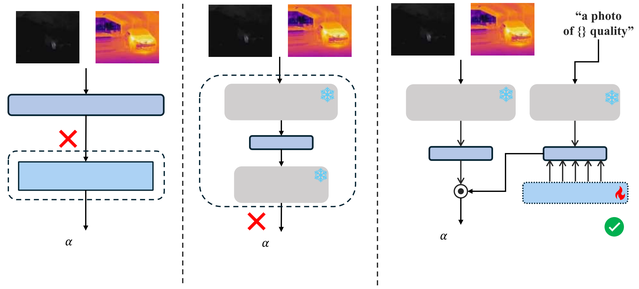


We propose the CoLA framework, which includes:

- 1) **Language-driven Quality Assessment** (LQA) to recalibrate image in put using a vision-language model, reducing noise impact without extra annotations,
- 2) **Conditional Dropout** (CD) to improve model performance when mo dalities are missing.

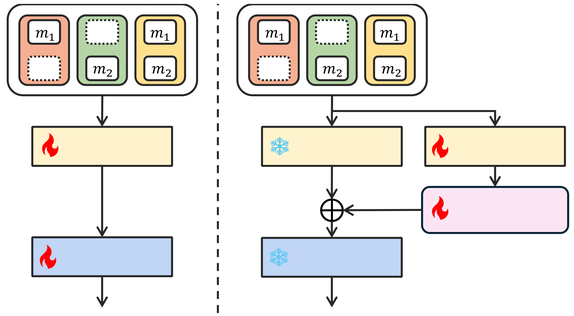
Language-driven Quality Assessment:

Architectural comparison of (a) No-Reference Method, (b) Pre-trained Assessment and (c) Our LQA.



Conditional Dropout:

Architectural comparison of (a) modality dropout and (b) Conditional Dropout in dual-modal object detection.



Results:

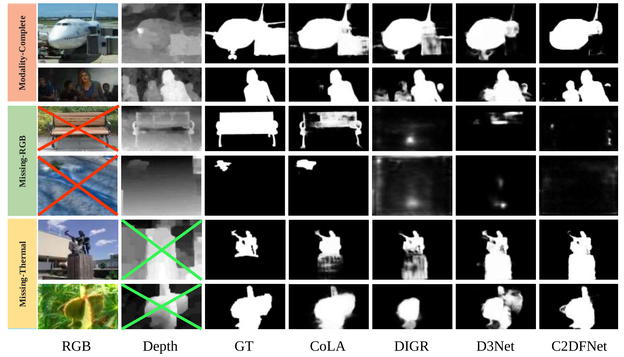
RGB-T SOD

Methods	VT821 Full	VT821 RGB	VT821 Thermal
DCNet (2022)	0.841	0.644	0.78
TAGFNet (2023)	0.825	0.727	0.771
LSNet (2023)	0.829	0.687	0.749
Ours	0.849	0.752	0.817

RGB-D SOD

Methods	DES Full	DES RGB	DES Depth
C^2DFNet (2022)	0.940	0.568	0.902
SPSN (2022)	0.950	0.793	0.908
HiDANet (2023)	0.979	0.907	0.922
Ours	0.963	0.947	0.926

Visualization:



Code Available

Wechat

WhatsApp